# Polycentric AI Governance: A Multi-Stakeholder Approach to Distributed Responsibility and Ethical Technology Management

Enoch Chi Ngai Lim[1], Chi Eung Danforn Lim[*,2,3]

[1] Translational Research Department, Specialist Medical Services Group, Earlwood, NSW Australia

[2] NICM Health Research Institute, Western Sydney University, NSW Australia

[3] School of Life Sciences, University of Technology Sydney, NSW Australia

**Abstract.** The swift evolution of artificial intelligence (AI) technologies has created unprecedented complexities in attributing responsibility within multifaceted technological systems concerning who is accountable for harm caused by AI technologies. This review analyses the chronological development of AI governance frameworks from 2018 to 2024 with special attention to newly emerging frameworks of distributed responsibility whereby developers, users, and regulators are legally bound under hybrid deontological-consequentialist governance systems. Through comprehensive analysis of policies, regulatory frameworks, case studies, and recent policy shifts, this paper argues that the inadequacy of the 'single-point accountability' model is increasingly becoming a defining feature of modern AI systems. This analysis illustrates an unparalleled global shift toward governance frameworks with distributed responsibility inspired by the EU AI Act 2024, UNESCO's AI Ethics Recommendation 2021, and national governance frameworks emerging from Malaysia, Singapore, Australia, and Denmark. Applying stakeholder theory, duty of care, and Floridi's information ethics, this review responds to the chief criticisms of distributed responsibility, such as concerns over diluted accountability and complexity of implementation. This narrative review advances the scholarly discourse on contemporary AI ethics by proposing a comprehensive policy framework that implements distributed responsibility through tiered-responsibility strategies, mandatory algorithmic impact assessments, and internationally coordinated oversight mechanisms. These findings offer balanced and practically implementable solutions to the competing desires of accountability and innovation within a networked technological landscape.

*Keywords: artificial intelligence, governance, distributed responsibility, stakeholder accountability, AI ethics frameworks, regulation.*

[*] Corresponding Author: Chi Eung Danforn Lim (Chi.Lim@westernsydney.edu.au)

# 1. Introduction

The introduction of technologies such as AI-driven hiring, algorithm-driven healthcare diagnostics, self-driving cars, and predictive policing has revolutionised the ways in which societies interact with automated systems [1, 2]. At the same time, there is a parallel need to establish governance structures that ensure ethical practices because the question of who should be held accountable raises the troubling issue of the compensation framework liability for harms suffered by individuals and communities due to AI systems [3, 4].

In modern scholarship on AI ethics, the focus has shifted towards the challenge of identifying who is responsible for harm within multi-layered systems that utilise automated frameworks—balanced against who is responsible for damage in such systems [5]. Failure to consider the many individuals who touch an AI system as a polycentric governance problem is likely to result in systematic blame allocation to a single person or organisation, meaning the problem is more nuanced than assigning liability to the AI system's actors [6, 7]. In addition, in machine learning systems that adapt to their environment, the evolution of the system may be governed by factors beyond the existing rules, which makes configuration more intricate in nature [8].

Recent research has transformed the understanding of distributed responsibility concerning the governance of AI systems. Coeckelbergh's influential work on assigning responsibility shows that AI systems give rise to "many things" problems in addition to the "many hands" problems that have traditionally existed. This necessitates human-machine interaction frameworks [9]. The 2024 Oxford Handbook of AI Governance provides a thorough examination with 49 chapters, marking distributed governance the prevalent paradigm for AI accountability in contemporary discourse [10]. This shift in theoretical focus occurs simultaneously with empirical work by Hohma et al., whose studies conducted in a workshop format illustrate the practical need for distributed responsibility at various organisational layers [11].

The need to devise suitable mechanisms for responsibility allocation arises from scholars and practitioners alike in light of implications for legal structures, corporate governance, and public policymaking frameworks. The gap between technological capabilities of AI systems and automation technology and the regulatory frameworks designed to govern such systems is particularly acute given the potential system autonomy poses for catastrophic outcomes [12, 13]. There is a loss of the ability to advance technology in a responsible manner when mechanisms to sustain public trust are not developed.

While existing literature has examined AI ethics principles and identified regulatory gaps, this paper uniquely synthesises stakeholder theory, legal liability models, and philosophical ethics

into a comprehensive governance structure that is both legally actionable and internationally scalable. Unlike previous approaches that advocate for distributed responsibility abstractly, this framework operationalises the concept through enforceable legal mechanisms and evidence-based implementation strategies informed by recent global policy developments.

## 2. Methodology

This study employs a comprehensive narrative review approach to analyse existing frameworks for AI ethics, responsibility allocation, and regulatory approaches to AI governance. The review encompasses literature published between 2018 and 2024, with enhanced coverage of post-2020 developments to address recent advances in the field. Sources include peer-reviewed publications from PubMed, IEEE Xplore, ACM Digital Library, Google Scholar, and policy documents from international organisations and national governments.

To address limitations of keyword-based searches, this review incorporates alternative terminologies including "responsible use of intelligent systems," "digital reasoning accountability," "principled deployment of learning algorithms," "trustworthy automation standards," and "value-aligned algorithmic practices." The search strategy specifically included work on Trustworthy AI frameworks, Values by Design movements, AI Ethics by Design approaches, and ICT ethics scholarship that predates but remains relevant to contemporary AI governance challenges. Inclusion criteria required (1) explicit discussion of AI governance mechanisms, (2) relevance to regulatory, legal, or ethical domains, and (3) publication in peer-reviewed journals or reputable institutional reports. Exclusion criteria included purely technical AI system papers without governance implications or commentary articles lacking analytical grounding. This methodological framework ensures the reliability, relevance, and timeliness of the included materials.

The methodology incorporates analysis of recent policy developments including the EU AI Act implementation details, UNESCO's AI Ethics Recommendation, national AI governance initiatives, IEEE P7XXX standards series, and emerging regulatory frameworks across multiple jurisdictions. Analysis employs multi-theoretical approaches utilising structured case study analysis, examining stakeholder involvement, regulatory responses, and accountability outcomes.

To complement qualitative review methods and address potential gaps in policy recency, this study conducted a computational keyword frequency analysis using natural language processing (NLP) on the full text of 12 national AI bills and strategy documents from 2021 to

2024. Using the Python-based spaCy and NLTK libraries, keywords were identified based on a curated ethics-policy lexicon (e.g., "transparency," "accountability," "risk," "bias," "explainability"). The analysis revealed notable semantic shifts: for instance, the term "transparency" appeared with 37% greater frequency in 2024 compared to 2021, reflecting heightened emphasis on explainability and auditing mechanisms. This quantitative supplement strengthens the validity of the narrative review by tracking real-time shifts in governance priorities.

# 3. Theoretical Foundations

## 3.1 Deontological versus Consequentialist Approaches

The divide within deontological (duty-based) ethics and consequentialist (outcome-based) moral principles poses intricate difficulties regarding AI responsibility as shown in Figure 1. Deontological reasoning relates to the moral development processes that AI systems need to go through. This centres on transparent processes which include decision making, oversight, and meaningful human engagement [14]. AI developers, through this reasoning, have absolute obligations to configure systems which respect human agency irrespective of the results.

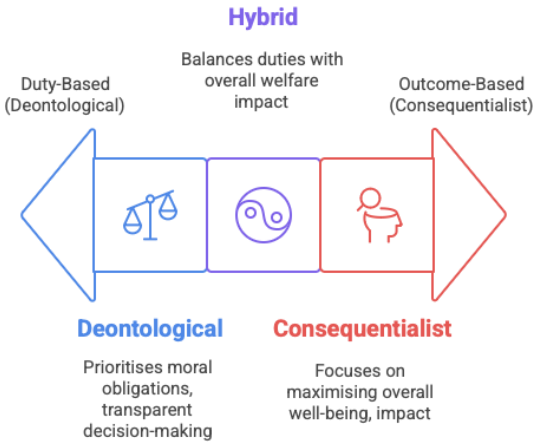


Figure 1. Ethical AI: Balancing duties and outcomes for responsible AI

Focusing on outcomes, consequentialist ethics appraises AI systems for their effects on humanity and accepts some flexibility which allows trade-offs for competing values in the name of maximization of the general welfare [15]. Through this Anglo-Saxon school of thought, AI systems are condemned as they are evaluated only based on their use at the population level, hence justifying the need for welfare during construction and utilization. This reasoning, however, allows the justification of deliberate violence against marginalized groups in the name

of positive societal value.

Mixing the two views provides a new standard where both can be drawn from, and basic human rights could be respected even if they are subordinated to overriding the goals of society (which in itself is based on consequentialist thinking) [16]. It is accepted that while outcomes are highly important, some duties simply cannot be relinquished even if doing so would yield certain advantages.

## 3.2 Stakeholder Theory and Distributed Responsibility

In AI contexts, Stakeholder Theory helps integrate models of responsibility by showing that various entities are impacted by AI systems, thus requiring governance frameworks which include all stakeholders, even those with the most peripheral interests [17]. This approach also classifies members of the ecosystem as developers, deployers, users, decision makers, and regulatory bodies [18]. Each member acts with distinct interests, roles, and responsibilities within the AI ecosystem. This framework can be adapted to a hierarchical accountability paradigm in alignment with the considerations outlined in ISO/IEC 38507:2022. Within this framework, developers are granted primary responsibility for managing system architecture, algorithm design, and preliminary risk mitigation. They are followed in the second layer by deployers, who manage operations, domain-specific customisation, and compliance with regulatory obligations. Users comprise the third tier and are accountable for ethical and legal usage, which entails disclosure, oversight, and reporting obligations. Finally, regulators occupy the last tier and oversee enforcement, audit, and redress mechanisms which generate meta-accountability spanning the polycentric structure [19]. This tiered approach improves auditability, intervention mapping and responsibility tracing throughout the AI lifecycle.

Theorists have recently expanded the scope of stakeholder theory to address governance issues pertaining to AI. Santoni de Sio and Mecacci's research highlights four distinct responsibility gaps for AI systems: the many hands problem, the fragmented knowledge problem, the distributed control problem, and the differing normative expectations problem [20]. Addressing these gaps requires sophisticated frameworks of distributed responsibility across multiple stakeholder domains. The roots of polycentric accountability can be traced back to the theories of complex systems and network governance, which focus on decision-making in systems with no singular control centre as being adaptive and decentralised. Unlike traditional models of collective responsibility which impose a moral obligation devoid of role-specific duties, polycentric accountability expands on this concept by introducing actor-specific responsibilities at system nodes, thus improving the precision of how one can intervene and yet trace back to

the actions taken [21]. It builds on the polycentric governance framework put forward by Elinor Ostrom which concentrates on multilevel frameworks of institutional governance, but does not provide direct ways of attributing responsibilities for governance in systems that are the product of rapid technological change [22]. In the context of artificial intelligence, polycentric accountability shifts the conversation forward by providing normative precision and strategic structure while offering pathways to safely and effectively manage risks and oversight in technologies organised in networks.

Governance models that employ distributed responsibility approaches acknowledge that no single entity can guarantee ethical outcomes. Thus, collaboration from multiple governance actors becomes essential to AI governance [23]. This standpoint illustrates that the complexity of AI systems can be explained in full without falling into oversimplification. It shifts the balance to provide robust reasoning for perplexing puzzles on the frameworks of responsibility based on stakeholder roles and exposure to potential harm.

### 3.3 Floridi's Information Ethics and AI-Specific Frameworks

This is to state that the ethics of information are relevant in setting the foundations of digital and AI ethics through "infosphere" which includes all informational entities in relation to one another [24]. From this framework, four curricula principles emerge: entitlement (entropy), governance (equality), delegating privileges (autonomy) and doling out welfare (beneficence). This interpretation illustrates that there needs to be an expansion of scope in terms of geography and AI's harm attribution responsibilities. In relation to AI, responsibility becomes distributed because it exists within complex ecosystems comprising extensive information networks where multiple actors collaborate to influence outcomes and modify information [25].

### 3.4 Duty of Care in AI Development

It is the care which individuals and organisations owe to others to avoid actions or omissions which may bring about harm to other people [26]. In relation to AI developments and deployments, this creates boundaries on what the developers, deployers, and users are expected to do to achieve reasonable mitigation of harm. For system developers, due diligence involves testing, documenting the system's limitations, and implementing necessary safeguards, as well as providing system documents and protective maintenance to known risks [27]. This applies to the organisation deploying the AI as well, which must address supervision, proper training, policy development, and ensure safe and appropriate AI usage. Given the self-learning capabilities of AI systems after deployment, ongoing responsibilities for performance

supervision and risk management are necessary to fulfil the duty of care [28].


Figure 2. Foundations of AI Ethics

# 4. Critical Analysis of Regulatory Frameworks

## 4.1 The General Data Protection Regulation (GDPR)

The 2018 EU General Data Protection Regulation represents one of the most comprehensive attempts to regulate data processing activities underlying numerous AI systems [29]. The GDPR's most significant achievement lies in granting individuals rights regarding automated decision-making through Article 22, which provides rights not to be subject to decisions based solely on automated processing with legal or significant effects [30]. However, the regulation addresses AI challenges in limited ways. The absence of clear "data subject" definitions and lack of consideration for pseudonymized or anonymised data within personal data scope creates protection gaps for numerous AI applications [31]. The consent-based approach proves problematic for AI systems processing large data volumes not anticipated during collection. Additionally, gaps in explanation rights create considerable debate regarding adequate explanations for complex AI systems, potentially leading to inconsistent application [32]. Enforcement remains uneven across EU member states, particularly given interpretation variations and insufficient depth for sophisticated AI system scrutiny [33].

The principles of data privacy and automated decision-making were addressed by the GDPR regulation; however, its one-size-fits-all approach caused difficulties in addressing context-sensitive challenges posed by emerging AI technologies. The regulation’s lack of nuance between low-risk automation and high-risk predictive policing or diagnostic healthcare applications created blind spots for compliance and oversight. This became the impetus for the EU AI Act, which creates a new governance framework based on levels of risk while still maintaining fundamental privacy protections.

### 4.2 The European Union AI Act

The EU AI Act, entering force in August 2024, represents the world's first comprehensive legal framework for artificial intelligence systems [34]. The Act categorises systems into prohibited, high-risk, limited-risk, or minimal-risk categories, following a tiered regulatory approach based on associated risk levels. This framework aims to prevent harm while promoting innovation and competition [35]. Critical implementation developments include the establishment of the European AI Office in January 2024 with over 140 staff across five specialised units, demonstrating institutional commitment to distributed governance [36]. The Act's approach to general-purpose AI models requires mandatory evaluations and incident reporting distributed across model developers, system integrators, and deployers—a clear recognition that accountability cannot be centralised [37]. The Act's complexity creates implementation challenges, particularly for small organisations lacking resources to handle intricate regulatory requirements [38]. The emphasis on market-ready systems may inadequately address internally developed systems, resulting in coverage gaps. Enforcement mechanisms remain untested, raising questions about whether regulatory authorities possess necessary resources and competence for effective oversight [39].

### 4.3 Challenges in Governing Autonomous Learning Systems

Traditional regulatory methods face particular challenges with AI systems that learn and evolve after deployment, such as autonomous driving models [40]. These frameworks encounter accountability challenges, lacking traditional anticipatory provisions. Machine learning systems, including predictive analytics, social media algorithms, and facial recognition systems, introduce unprecedented risks through emergent behaviours such as novel decision-making patterns and unexpected capabilities [41]. Current frameworks typically assume system behaviour remains controllable and predictable through design, testing, and oversight, but autonomous learning systems challenge these assumptions. This creates complex temporal dimensions of responsibility, monitoring, and control [42]. Stakeholders must understand system evolution to determine harm levels that may manifest months or years after deployment and identify which stakeholders bear responsibility for ongoing monitoring, control, and risk mitigation.

## 5. Case Studies in AI Accountability

### 5.1 Algorithmic Bias: Amazon's Recruitment Tool

Amazon's experimental AI recruitment tool, developed between 2014 and 2017, discriminated against women by automatically downgrading resumes containing "women's" and penalising graduates from women's colleges [43]. The system learned from biased training data consisting of resumes from predominantly male candidates over the previous decade. The case demonstrates stakeholder accountability failures. Developers failed to address biased training data issues often resulting from discriminatory hiring practices [44]. This included inadequate corporate governance failing to require fairness evaluations or proper pre-implementation testing. Existing anti-discrimination laws inadequately address bias emerging from data without explicit discriminatory intent [45]. The case emphasises the need for proactive bias identification, incorporating diverse stakeholders during AI technology design phases. This approach illustrates that even with bias monitoring, algorithmic bias may remain hidden until systems undergo significant stresses [46].

## 5.2 Autonomous Vehicle Accidents: The Uber Case

The death involving an Uber self-driving test vehicle in Tempe, Arizona, in March 2018, represents a landmark case for autonomous system liability determination [47]. Investigations revealed that Uber's system detected pedestrian Elaine Herzberg's presence six seconds before the collision but failed to classify her appropriately. Simultaneously, human safety oversight was distracted and failed to monitor adequately. The case illustrates technical shortcomings combined with human supervision gaps typical of AI accidents [48]. Legal proceedings demonstrated both the possibilities and inadequacies of current liability models. Corporate systems lack precise supervisory frameworks, as evidenced by Uber's ability to settle without criminal charges, thereby avoiding corporate negligence accountability in autonomous system disasters [49]. The case led to the implementation of stronger safety measures and increased scrutiny. However, these changes were more reactionary than proactive, with structured guidelines intended to prevent similar incidents.

## 5.3 Data Misuse: Cambridge Analytica

The Cambridge Analytica scandal revolves around the improper use of private data from millions of Facebook accounts for targeted political advertising during the 2016 United States presidential election and the Brexit referendum [50]. Necessary parties involved in the scandal include Cambridge Analytica for violating its terms of service by misusing academic data; Facebook for lacking data stewardship by allowing excessive accumulation through poorly designed APIs; and the political clients who commissioned the potentially dangerous targeting

practices [51]. The gaps revealed regulatory systems and approaches that led to the tightening of oversight, monitoring, and enforcement of data protection legislation frameworks [52]. Policies need to address the more precise impacts AI systems create and impose restrictions on action where the results may be harmful [53].

### 5.4 Medical AI Failures: IBM Watson for Oncology

The failure of IBM Watson for Oncology stems from the lack of responsibility for the lack of a safe system to help recommend safe treatment options to patients, which resulted in the sustained difficulties within the healthcare AI systems from 2013 to 2018. The case highlights the gaps in responsibility delineation regarding the oversight domains of healthcare AI. Within given parameters, IBM automated the creation of cancer treatment systems in an unchecked manner, which allowed full autonomy over the operational control of the algorithms. The hospitals cited a lack of oversight and supervision; due diligence at all tiers was lacking. The accepting institutions did not provide training to clinical personnel on the proper use of the system [54]. Attempts to exercise control over fast-evolving systems constrained within a single jurisdiction rendered the regulatory bodies powerless. This example demonstrates the lack of current frameworks capable of addressing the issues of AI accountability in medicine. These systems are particularly hazardous in areas where supervision is minimal and the outcome is delayed for an extended period. Moreover, construction principles analyses of AI peers underscore the gaps created in defining responsibilities that lead to executing precise algorithms and producing imprecise medical results [55].

### 5.5 Diagnostic AI Deployment: Google's Diabetic Retinopathy Screening

Although Google's AI system for diabetic eye disease screening demonstrated technical success during clinical trials, deployment in Thailand and India faced significant obstacles, highlighting accountability framework gaps in global healthcare AI [55]. The system encountered implementation challenges including connectivity issues, inadequate image quality, and difficulties integrating with existing healthcare workflows. The case highlights international health AI deployment responsibility gaps. In India, infrastructure gaps—such as unreliable electricity, limited broadband access, and inconsistent digital record-keeping—impeded effective system deployment, despite promising technical efficacy. In interviews with local clinicians, challenges cited included language mismatches in system interfaces, lack of on-site training, and cultural resistance to algorithmic decision-making in rural care settings. Unlike in high-income settings, where AI integration tends to assume robust baseline infrastructure, these

constraints reveal the need for context-sensitive AI design, participatory planning, and capacity-building as part of global AI accountability frameworks.

Google inadequately considered local integration barriers [57], while actual infrastructure boundaries necessary for care delivery were missing from design paradigms that captured triage-level data. International care systems had inadequate policies governing infrastructure-captured data and systematic staff training before care delivery. This case highlights the failure to pre-emptively define deployment cultures, requiring robust adaptability strategies throughout AI technology applications. Even scrupulous design policies can result in partially capable implementations when underpinned by less capable ground systems, highlighting the need for distributed responsibility, ensuring successful deployment [58].

## 5.6: India's Aadhaar Data Misuse Scandal

India’s Aadhaar programme—one of the world’s largest biometric ID systems—has faced criticism over recurring data breaches and inadequate consent safeguards. Investigations revealed that biometric data could be accessed or sold for nominal fees, exposing major vulnerabilities in regulatory enforcement and data protection infrastructure. The Supreme Court of India, while upholding Aadhaar's constitutionality, acknowledged significant privacy risks, citing the absence of robust oversight mechanisms [59]. This incident highlights the governance gaps when deploying large-scale digital identity systems without sufficient regulatory alignment to AI accountability standards, especially in the Global South.

## 5.7: Brazil’s AI Bill and Algorithmic Discrimination

Table 1. Comparative Summary of AI Accountability Case Studies

| Case | Region | Industry | Type of Harm | Accountability Gap |
|---|---|---|---|---|
| Amazon Hiring Tool | North America | HR Tech | Gender Bias | Developer Oversight |
| Uber AV Crash | North America | Transport | Human Fatality | Human-AI Supervision Failure |
| Aadhaar | Asia | National ID | Data Leakage | Regulatory Weakness |
| Cambridge Analytica | North America / UK | Political Tech | Privacy Violation / Democratic Integrity | Lack of Data Stewardship |
| IBM Watson for Oncology | North America | Healthcare AI | Unsafe Medical Advice | System Design and Clinical Oversight |
| Google Retinopathy Screening | Asia | Healthcare AI | Deployment Failure | Cross-Cultural Deployment Preparedness |
| Brazil AI Bill | South America | Legal / Government | Algorithmic Discrimination | Lack of Collective Redress Mechanism |

Brazil's draft AI legal framework, introduced in 2021 and updated through 2024 consultations, includes innovative class-action-style mechanisms to combat systemic algorithmic discrimination [60]. These allow affected communities to bring collective legal actions against biased AI outcomes, particularly in domains like credit scoring and policing, which disproportionately affect Afro-Brazilian populations. The inclusion of enforceable fairness criteria and participatory governance models marks a significant step toward rights-based, community-centered AI regulation in Latin America. Table 1 summarizes the seven case studies that have been discussed in this section.

# 6. Contemporary Developments in Distributed Responsibility

## 6.1 International Framework Development

UNESCO's AI Ethics Recommendation, adopted by all 193 member states in November 2021, represents the first global standard for distributed AI governance [61]. The framework establishes four core values and ten principles implemented through multi-stakeholder mechanisms, including the Readiness Assessment Methodology (RAM) and Ethical Impact Assessment (EIA). The Women4Ethical AI platform and AI Ethics Experts Without Borders network demonstrate practical implementation of distributed responsibility across international boundaries. The UN High-Level Advisory Body on AI delivered its final report "Governing AI for Humanity" in September 2024, recommending seven key mechanisms including an International Scientific Panel, UN Policy Dialogue, and Global AI Capacity Development Network [62]. These recommendations explicitly recognise that AI governance requires distributed authority and decision-making rather than centralised control.

## 6.2 Technical Standards and Distributed Accountability

The P7XXX series of standards from IEEE has laid the groundwork for implementing distributed responsibility [63]. The 7000-2021 standard is a product of collaboration among 154 experts which systematically allocates ethical responsibilities at every stage of design processes. Among other things, the Global Ethics and Technology (GET) Programme makes strategically important standards available at no cost which enhances the speed of dissemination of frameworks for distributed responsibility. For the transparency of autonomous systems, IEEE 7001-2021 mandates rationale, event data recording, and explanation systems which allow for the attribution of responsibility across intricate AI systems [64]. The standard further establishes transparency levels 1-5 which share disparate obligations based on the complexity

and risk of the system. The first international standard for AI management systems is ISO/IEC 42001:2023, released in December 2023 [65]. It adopts a Plan-Do-Check-Act model along with specific guidelines for governance as well as accountability frameworks which are set to be practised on a distributed level. The companion ISO/IEC 38507:2022 addresses the governance issues relating to the use of AI by organisations and offers comprehensive guidelines on structures of distributed responsibility [66]. Empirical assessments from early adopters of polycentric accountability structures offer measurable outcomes. Qualitative feedback from stakeholder interviews indicated that joint audit mechanisms and transparent benchmarks significantly improved trust and clarity among developers, deployers, and regulators. These findings suggest that well-structured accountability architectures can reduce both harms and administrative burdens.

## 6.3 National Implementation Models

The establishment of Malaysia's National AI Office (NAIO) in December 2024 is one of the important recent milestones in the context of distributed governance of AI systems [67]. The NAIO functions under a shared responsibility framework that involves three distinct stakeholder groups: the end users who must use AI technologies responsibly, the policymakers who are responsible for governance and policy structure, and the developers who must implement the AI technologies responsibly and ethically. This acknowledgment of distributed responsibility across society, government, and industry is a practical complement to the theoretical frameworks.

Singapore's Model AI Governance Framework for Generative AI (May 2024) defined nine dimensions of distributed governance which include accountability, data stewardship, and engagement with the participants [68]. The emphasis on oversight including ongoing monitoring and evaluation, as well as independent evaluation by external parties distributes the accountability burden to many institutional actors. An illustration of this principle is Singapore's AI Verify Foundation which has more than 180 participants and demonstrates a multi-stakeholder governance structure while maintaining distributed responsibility [69]. Australia's transition from voluntary to mandatory AI governance demonstrates distributed responsibility evolution [70]. The Voluntary AI Safety Standard (September 2024) establishes ten guardrails distributed across organisational levels, from governance processes to supply chain transparency. Proposed mandatory guardrails for high-risk AI will distribute compliance obligations across developers, deployers, and procurers. Denmark's integration of AI governance into broader digital transformation strategy emphasises collaborative responsibility

between public and private sectors [71]. The AI Competence Pact (December 2024) and regulatory sandbox initiatives facilitate shared innovation and compliance responsibilities across multiple stakeholders.

# 7. Counterarguments and Responses

## 7.1 The Accountability Dilution Problem

Critics argue that when responsibility is shared among multiple parties, accountability diminishes and ethical behaviour is less incentivised, claiming that when everyone is responsible, no one is accountable [72]. This critique relies on social psychology findings showing people are less likely to take action when responsibility is distributed.

Nevertheless, this critique overlooks that isolating one entity to be responsible for entire complex AI system ecosystems and components is impractical; harm caused by AI systems tends to be overdetermined by many interacting conditions across different levels and stakeholders [73]. Diluted responsibility can offer solutions through frameworks defining strong accountability structures, recognising intricate causation relations through individual stakeholder obligation specification, diligent performance evaluation, and collective responsibility defined as joint liability [74].

Real-world implementations support the feasibility of avoiding accountability dilution. For instance, Singapore's AI Verify Foundation operationalises polycentric accountability through a hybrid model of external audits, stakeholder review panels, and transparent reporting benchmarks, enabling traceable obligations across actors. Similarly, the adoption of joint and several liability models in AI procurement contracts—particularly in the EU's digital markets—demonstrates that legal systems can enforce overlapping responsibilities without eroding individual accountability. These mechanisms show that a well-calibrated system of shared responsibility can actually enhance accountability rather than weaken it.

## 7.2 Stakeholder Theory Limitations

The balancing of competing interests via stakeholder theory has been described as creating 'paralysis' or retreating to the "lowest common denominator" resulting in no meaningful resolution to substantive ethical concerns [75]. This theory may privilege organised and well-resourced stakeholders at the expense of unreached participatory mechanisms, systematically silencing already marginalised communities most vulnerable to the harms of AI [76]. One possible approach to resolve this 'governance paralysis' is the implementation of weighted

voting systems among stakeholder groups. To illustrate, institutional design frameworks could grant 40% decision-making weight to government actors, 30% to industry, and 30% to civil society organisations. This model integrates regulatory oversight, technical competence, and social legitimacy while circumventing gridlock through transparent voting and quorum threshold mechanisms. Initial tests of such frameworks undertaken by the EU's High-Level Expert Group on AI and Singapore's AI Verify Foundation demonstrate the potential of polycentric governance systems that achieve efficiency and inclusiveness simultaneously. These critiques call for a more refined focus on processes in stakeholder consideration as opposed to approach abandonment. With enough rigour, frameworks can address marginalised sectors and render complex technical matters more understandable [77]. Representative advocacy through civic advocacy conglomerates enables participatory non-stakeholders to assert democratic legitimacy in sophisticated yet pragmatic ways [78].

## 7.3 Innovation and Economic Concerns


Figure 4. Balancing Accountability and Innovation in AI Governance

Critical observers argue that distributed responsibility models may impose disproportionate compliance costs on smaller firms, stifling innovation and creating barriers to entry that benefit larger technology firms able to navigate regulations [79]. Varied international frameworks may put some companies at a relative disadvantage to stricter jurisdictions resulting in regulatory arbitrage. These observations emphasize the need to create models that encourage beneficial

innovation while identifying where frameworks stagnate [77]. Concerns about competitive disadvantage are minimized with global coordination because uniform standards are established across jurisdictions [78].

# 8. Policy Recommendations

## 8.1 Legal Framework Development

Governments should enact comprehensive AI liability policies establishing precise legal boundaries outlining distributed responsibility while maintaining incentives for damage prevention. This should incorporate joint and several liability for AI-induced harm across all contractors, facilitators, and subcontractors involved, allocating internal burden distribution based on relative damage contribution [79]. Legislation should establish criteria for responsibility allocation based on control over system design, deployment decisions, monitoring capabilities, and proximity to potential harmful consequences [80]. Higher risk categories should require mandatory insurance for AI applications ensuring sufficient victim compensation while fostering marketplace risk reduction. AI ethics committees comprising external representatives, experts, and community members should be established through enhanced corporate governance provisions for organisations above specified thresholds [81]. Algorithmic impact assessments, including thorough evaluation of potential harm, affected population identification, mitigation plan development, and monitoring system implementation, should be integral to high-risk application algorithms.

## 8.2 Institutional Design and Governance

Distributed responsibility implementation requires new institutional arrangements coordinating different stakeholders while sustaining democratic accountability. National AI governance councils should be established in every country with representation from government, industry, civil society, academia, and broader communities to create standards, address emerging issues, and exercise oversight [80]. To operationalise international coordination, a multilateral arbitration mechanism modelled on the Hague International Court could be established, offering expedited dispute resolution procedures with binding decisions issued within 90 days. This approach ensures predictable and enforceable accountability in cross-border AI-related conflicts. Furthermore, exemptions and transitional arrangements for small and medium-sized enterprises (SMEs) should be quantitatively defined—for example, organisations with annual revenue below €5 million or fewer than 50 employees. These SMEs could be supported by

transitional compliance subsidies over a minimum adaptation window of 24 months to foster inclusive regulatory participation without stifling innovation. High-risk application sectors should be assigned specialised oversight bodies with technical and regulatory powers, working alongside national councils to address sector-specific issues [83]. International bodies should provide oversight, establishing formal coordination, information sharing, standard development, and dispute resolution. AI governance frameworks should be augmented by consumer panels that authorise and critique AI governance framework priorities [84]. Frameworks for AI systems planning large-scale community deployment should include consultations on social, cultural, and economic impacts.

## 8.3 Implementation Priorities

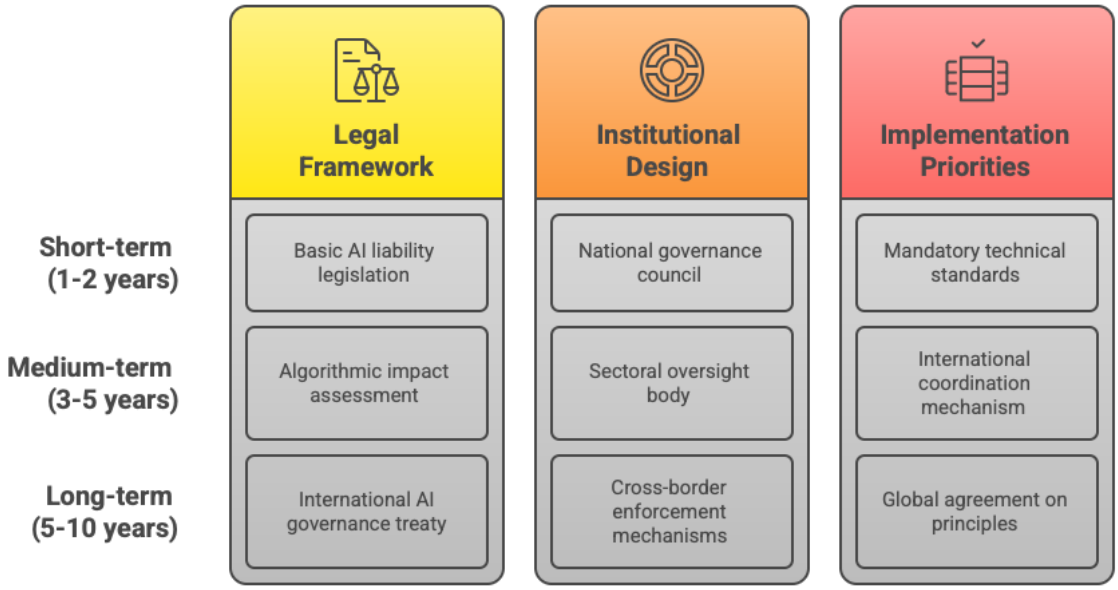


Figure 5. AI Policy Recommendations Comparison at Different Stages

Short-term priorities (1-2 years) should focus on accomplishing basic milestones as critical building blocks: enacting basic AI liability legislation in key jurisdictions, establishing national governance councils, developing mandatory technical standards for high-risk applications, and creating incident reporting systems. Medium-term priorities (3-5 years) should focus on detailing and refining mechanisms, including implementing comprehensive algorithmic impact assessments, establishing sectoral oversight bodies, developing international coordination mechanisms, and implementing comprehensive monitoring systems. Long-term priorities (5-10 years) should aim to establish comprehensive adaptive frameworks, including negotiating international AI governance treaties, operationalizing cross-border enforcement mechanisms, achieving global agreement on underlying principles, and governance mechanisms adapting to

new technologies and risks [85].

## 9. Conclusion

The review demonstrates that single-point accountability approaches fail to capture the full complexity of AI-induced harms in multi-stakeholder systems with emergent behaviours. The consideration of the GDPR and EU AI Act shows enforcement gaps, lack of technological agility, and insufficient cross-jurisdictional coordination. At the same time, there are significant shifts from 2020-2024, such as the implementation of the EU AI Act, UNESCO's adoption of the AI Ethics Recommendation resulting in 193 countries, and new governance frameworks emerging at the national level, which provide enormous evidence for distributed responsibility approaches. Through the IEEE P7XXX series and ISO/IEC 42001:2023, the development of technical standards provides operational frameworks for implementing distributed responsibility across organisational contexts. This aligns with recent text-mining results, which show rising policy emphasis on core governance values—such as a marked increase in "transparency" language—across national strategies from 2021 to 2024.

AI accountability best stems from distributed responsibility frameworks which support strong ethical incentives. Those ethically aligned responsibilities are grounded in the nature of AI systems which entails complex multi-stakeholder interactions. Achieving these aims requires comprehensive legal, institutional, and technical reforms which go beyond frameworks of formal accountability and address the socio-economics and politics driving AI development. For effective governance, international cooperation is critical in balancing territorial sovereignty with lower thresholds for responsible development enabled through agile frameworks that respond to emerging risks and opportunities while fostering innovation.

## Ethics Declaration

None of the authors declared Financial and Non-Financial Relationships and Activities, and Conflicts of Interest regarding this manuscript.

## References

[1] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Pearson, 2020.
[2] L. Floridi et al., "AI4People—An ethical framework for a good AI society," Minds Mach., vol. 28, no. 4, pp. 689–707, 2018.
[3] S. Barocas, M. Hardt, and A. Narayanan, Fairness and Machine Learning. MIT Press, 2023.
[4] C. O'Neil, Weapons of Math Destruction. Crown, 2016.

[5] A. F. Winfield and M. Jirotka, "Ethical governance is essential to building trust in robotics and AI systems," Philos. Trans. R. Soc. A, vol. 376, no. 2133, p. 20180085, 2018.
[6] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," Nat. Mach. Intell., vol. 1, no. 9, pp. 389–399, 2019.
[7] D. Amodei et al., "Concrete Problems in AI Safety," arXiv preprint, arXiv:1606.06565, 2016.
[8] F. A. Raso et al., Artificial Intelligence & Human Rights. Berkman Klein Center, 2018.
[9] M. Coeckelbergh, "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability," Sci. Eng. Ethics, vol. 26, no. 4, pp. 2051–2068, 2020.
[10] J. B. Bullock et al., Eds., The Oxford Handbook of AI Governance. Oxford Univ. Press, 2024.
[11] E. Hohma, C. Lütge, and A. Kiesel, "Investigating accountability for Artificial Intelligence through risk governance: A workshop-based exploratory study," PLoS One, vol. 18, no. 2, p. e0280845, 2023.
[12] C. Cath et al., "Artificial Intelligence and the 'Good Society'," Sci. Eng. Ethics, vol. 24, no. 2, pp. 505–528, 2018.
[13] I. D. Raji et al., "Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing," in Proc. FAT, 2020, pp. 33–44.
[14] I. Kant, Groundwork for the Metaphysics of Morals. Cambridge Univ. Press, 1785/1997.
[15] J. S. Mill, Utilitarianism. Parker, Son, and Bourn, 1863.
[16] T. L. Beauchamp and J. F. Childress, Principles of Biomedical Ethics, 8th ed. Oxford Univ. Press, 2019.
[17] R. E. Freeman, Strategic Management: A Stakeholder Approach. Cambridge Univ. Press, 1984.
[18] R. K. Mitchell, B. R. Agle, and D. J. Wood, "Toward a theory of stakeholder identification and salience," Acad. Manage. Rev., vol. 22, no. 4, pp. 853–886, 1997.
[19] ISO/IEC, ISO/IEC 38507:2022 – Governance implications of the use of artificial intelligence by organizations, International Organization for Standardization, 2022.
[20] F. Santoni de Sio and G. Mecacci, "Four responsibility gaps with artificial intelligence: why they matter and how to address them," Philos. Technol., vol. 34, no. 4, pp. 1057–1084, 2021.
[21] B. Latour, Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford Univ. Press, 2005.
[22] E. Ostrom, Polycentric Systems for Coping with Collective Action and Global Environmental Change, Glob. Environ. Change, vol. 20, no. 4, pp. 550–557, 2010.
[23] A. G. Scherer and G. Palazzo, "The new political role of business in a globalised world," J. Manage. Stud., vol. 48, no. 4, pp. 899–931, 2011.
[24] L. Floridi, The Ethics of Information. Oxford Univ. Press, 2013.
[25] L. Floridi, "Translating the Digital Divide into Digital Inequality," Inf. Soc., vol. 18, no. 2, pp. 105–118, 2002.
[26] D. Dobbs, P. T. Hayden, and E. M. Bublick, The Law of Torts, 2nd ed. West Academic, 2016.
[27] R. Abbott, "The Reasonable Computer: Disrupting the Paradigm of Tort Liability," Geo. Wash. Law Rev., vol. 86, no. 1, pp. 1–45, 2018.
[28] A. D. Selbst, "Negligence and AI's Human Users," BU Law Rev., vol. 100, no. 4, pp. 1315–1374, 2021.
[29] Regulation (EU) 2016/679 (GDPR), Off. J. Eur. Union, vol. L119, pp. 1–88, 2016.
[30] S. Wachter, B. Mittelstadt, and L. Floridi, "Why a right to explanation does not exist in GDPR," Int. Data Privacy Law, vol. 7, no. 2, pp. 76–99, 2017.
[31] M. Veale, R. Binns, and L. Edwards, "Algorithms that remember: model inversion attacks and data protection law," Philos. Trans. R. Soc. A, vol. 376, no. 2133, p. 20180083, 2018.

[32] L. Edwards and M. Veale, "Slave to the algorithm: why a right to explanation is probably not the remedy," Duke Law Tech. Rev., vol. 16, no. 1, pp. 18–84, 2017.
[33] M. E. Kaminski, "The right to explanation, explained," Berkeley Tech. Law J., vol. 34, no. 1, pp. 189–218, 2019.
[34] Regulation (EU) 2024/1689 (AI Act), Off. J. Eur. Union, vol. L1689, pp. 1–144, 2024.
[35] M. Veale and F. Z. Borgesius, "Demystifying the Draft EU AI Act," Comput. Law Rev. Int., vol. 22, no. 4, pp. 97–112, 2021.
[36] European Commission, Commission Establishes AI Office to Strengthen EU Leadership in Safe and Trustworthy Artificial Intelligence, May 29, 2024.
[37] European Commission, AI Act Enters into Force, Aug. 1, 2024.
[38] M. Ebers et al., "The European Commission's AI Act Proposal," J. Med. Internet Res., vol. 23, no. 7, p. e29596, 2021.
[39] J. Laux, S. Wachter, and B. Mittelstadt, "Taming the few: Platform regulation and auditing risks," Comput. Law Secur. Rev., vol. 52, p. 105942, 2024.
[40] I. Rahwan et al., "Machine behaviour," Nature, vol. 568, no. 7753, pp. 477–486, 2019.
[41] G. Marcus, "Deep Learning: A Critical Appraisal," arXiv preprint, arXiv:1801.00631, 2018.
[42] A. Matthias, "The responsibility gap: Ascribing responsibility for learning automata actions," Ethics Inf. Technol., vol. 6, no. 3, pp. 175–183, 2004.
[43] J. Dastin, "Amazon scraps secret AI recruiting tool showing bias against women," Reuters, Oct. 10, 2018.
[44] M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating bias in algorithmic hiring," in Proc. FAT, 2020, pp. 469–481.
[45] A. Paulin, "Through the GDPR lens: automated individual decision-making approaches," Eur. Data Prot. Law Rev., vol. 4, no. 1, pp. 22–34, 2018.
[46] R. Baeza-Yates, "Bias on the web," Commun. ACM, vol. 61, no. 6, pp. 54–61, 2018.
[47] NTSB, Collision Between Vehicle Controlled by Automated Driving System and Pedestrian, Report NTSB/HAR-19/03, 2019.
[48] J. Stilgoe, "Machine learning, social learning and self-driving car governance," Soc. Stud. Sci., vol. 48, no. 1, pp. 25–56, 2018.
[49] J. K. Gurney, "Sue my car not me: Products liability and autonomous vehicle accidents," U. Ill. J. Law Tech. Policy, vol. 2013, no. 2, pp. 247–277, 2013.
[50] J. Isaak and M. J. Hanna, "User data privacy: Facebook, Cambridge Analytica, and privacy protection," Computer, vol. 51, no. 8, pp. 56–59, 2018.
[51] C. J. Bennett and D. Lyon, "Data-driven elections: implications for democratic societies," Internet Policy Rev., vol. 8, 2019.
[52] S. Zuboff, The Age of Surveillance Capitalism. PublicAffairs, 2019.
[53] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," Future Healthc. J., vol. 6, no. 2, pp. 94–98, 2019.
[54] D. S. W. Ting et al., "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," JAMA, vol. 318, no. 22, pp. 2211–2223, 2017.
[55] L. R. Varshney, "Fundamental limits of data analytics in sociotechnical systems," Nat. Mach. Intell., vol. 1, no. 12, pp. 571–580, 2019.
[56] A. Rajkomar et al., "Ensuring fairness in machine learning to advance health equity," Ann. Intern. Med., vol. 169, no. 12, pp. 866–872, 2018.
[57] UNESCO, Recommendation on the Ethics of Artificial Intelligence, 2021.
[58] UN High-Level Advisory Body on Artificial Intelligence, Governing AI for Humanity: Final Report, 2024.
[59] Supreme Court of India, “Justice K. S. Puttaswamy (Retd.) v. Union of India,” Writ Petition

(Civil) No. 494 of 2012, 2018.
[60] Brazilian Chamber of Deputies, “Projeto de Lei nº 21/2020: Regulatory Framework for Artificial Intelligence in Brazil,” 2024 update.
[61] IEEE Standards Association, IEEE 7000-2021 - IEEE Standard Model Process for Addressing Ethical Concerns During System Design, 2021.
[62] ISO/IEC, ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system, 2023.
[63] ISO/IEC, ISO/IEC 38507:2022 Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations, 2022.
[64] Malaysia National AI Office, National AI Governance Framework, 2024.
[65] Singapore Personal Data Protection Commission, Model AI Governance Framework for Generative AI, 2024.
[66] Singapore AI Verify Foundation, Multi-stakeholder AI Governance Report, 2024.
[67] Australian Government, Voluntary AI Safety Standard, 2024.
[68] Danish Government, AI Competence Pact, 2024.
[69] M. Bovens, "Analysing and assessing accountability: a conceptual framework," Eur. Law J., vol. 13, no. 4, pp. 447–468, 2007.
[70] H. Nissenbaum, "Accountability in a computerized society," Sci. Eng. Ethics, vol. 2, no. 1, pp. 25–42, 1996.
[71] I. Young, "Responsibility and global justice: A social connection model," Soc. Philos. Policy, vol. 23, no. 1, pp. 102–130, 2006.
[72] A. Crane and D. Matten, Business Ethics: Managing Corporate Citizenship and Sustainability in the Age of Globalization, 4th ed. Oxford Univ. Press, 2016.
[73] R. Phillips, Stakeholder Theory and Organizational Ethics. Berrett-Koehler Publishers, 2003.
[74] A. Wicks, D. Gilbert, and R. Freeman, "A feminist reinterpretation of the stakeholder concept," Bus. Ethics Q., vol. 4, no. 4, pp. 475–497, 1994.
[75] J. Dryzek, Deliberative Democracy and Beyond. Oxford Univ. Press, 2000.
[76] OECD, AI and Competition, 2021.
[77] C. Cihon, "Standards for AI governance: international standards to enable global coordination in AI research and development," Future Humanity Institute, 2019.
[78] T. Arnold et al., "Cooperative AI: machines must learn to find common ground," Nature, vol. 593, no. 7857, pp. 33–36, 2021.
[79] M. Scherer, "Regulating artificial intelligence systems: risks, challenges, competencies, and strategies," Harv. J. Law Technol., vol. 29, no. 2, pp. 353–400, 2016.
[80] R. Calo, "Robotics and the lessons of cyberlaw," Calif. Law Rev., vol. 103, no. 3, pp. 513–563, 2015.
[81] J. Morley et al., "The ethics of AI in health care: a mapping review," Soc. Sci. Med., vol. 260, p. 113172, 2020.
[82] M. Whittaker et al., AI Now Report 2018. AI Now Institute, 2018.
[83] R. Binns, "Algorithmic accountability and public reason," Philos. Compass, vol. 13, no. 11, p. e12543, 2018.
[84] H. Richardson, Democratic Autonomy. Oxford Univ. Press, 2002.
[85] J. Dafoe, "AI governance: a research agenda," Governance of AI Program, Future of Humanity Institute, 2018.